



HAL
open science

GSAAn: An alternative to gene set enrichment analysis

Aarón Ayllón-Benítez, Patricia Thébault, Fleur Mougin

► **To cite this version:**

Aarón Ayllón-Benítez, Patricia Thébault, Fleur Mougin. GSAAn: An alternative to gene set enrichment analysis. Journées Ouvertes de Biologie Informatique et Matématiques, Jul 2019, Nantes, France. hal-02401873

HAL Id: hal-02401873

<https://hal.science/hal-02401873>

Submitted on 10 Dec 2019

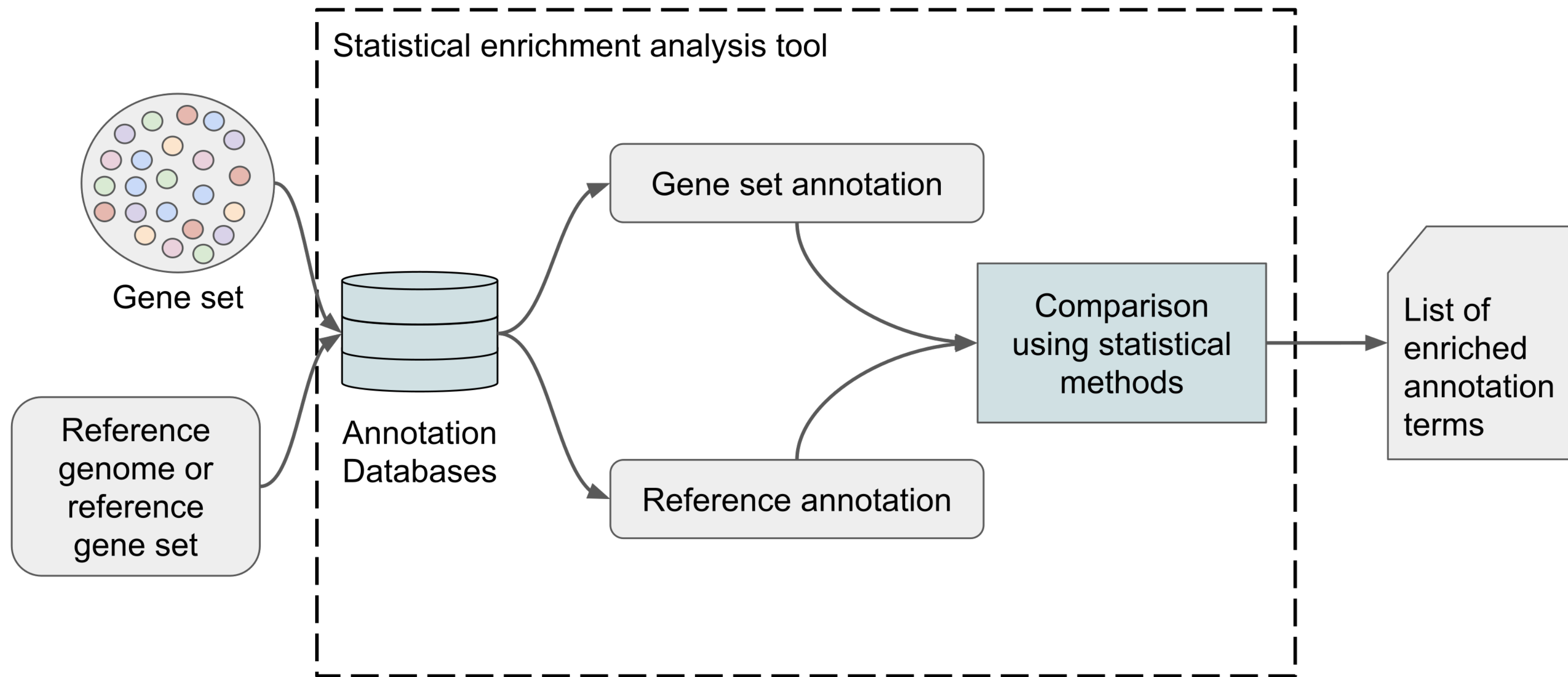
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 - Introduction

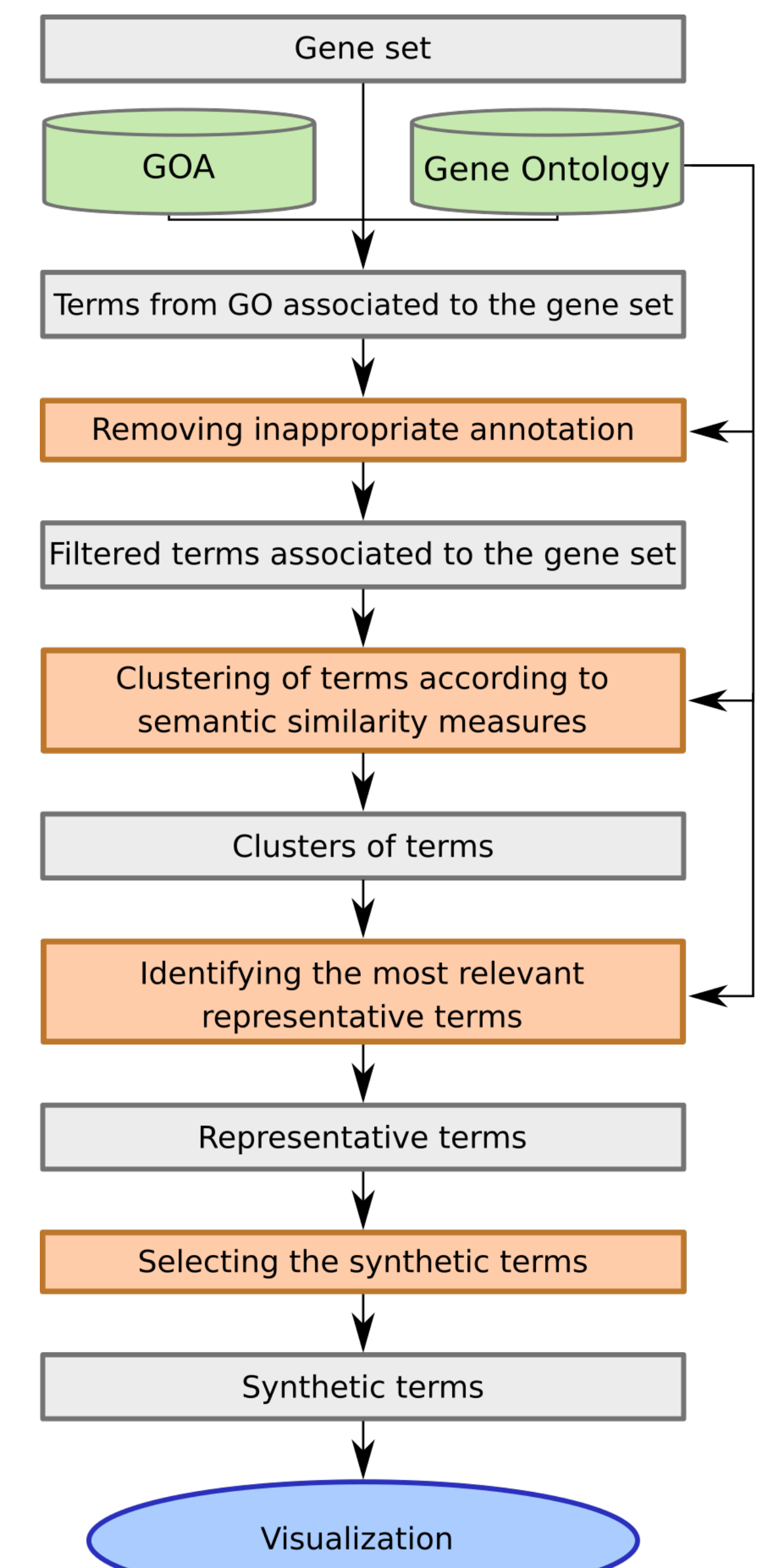
- The revolution in new sequencing technologies has strongly supported the production of omics data to improve our understanding of the relations between genotype and phenotype.
- Analyzing gene sets to identify their biological function and to synthesize the key annotation information aims to help biologists in their interpretation.

2 - Statistical enrichment limitations



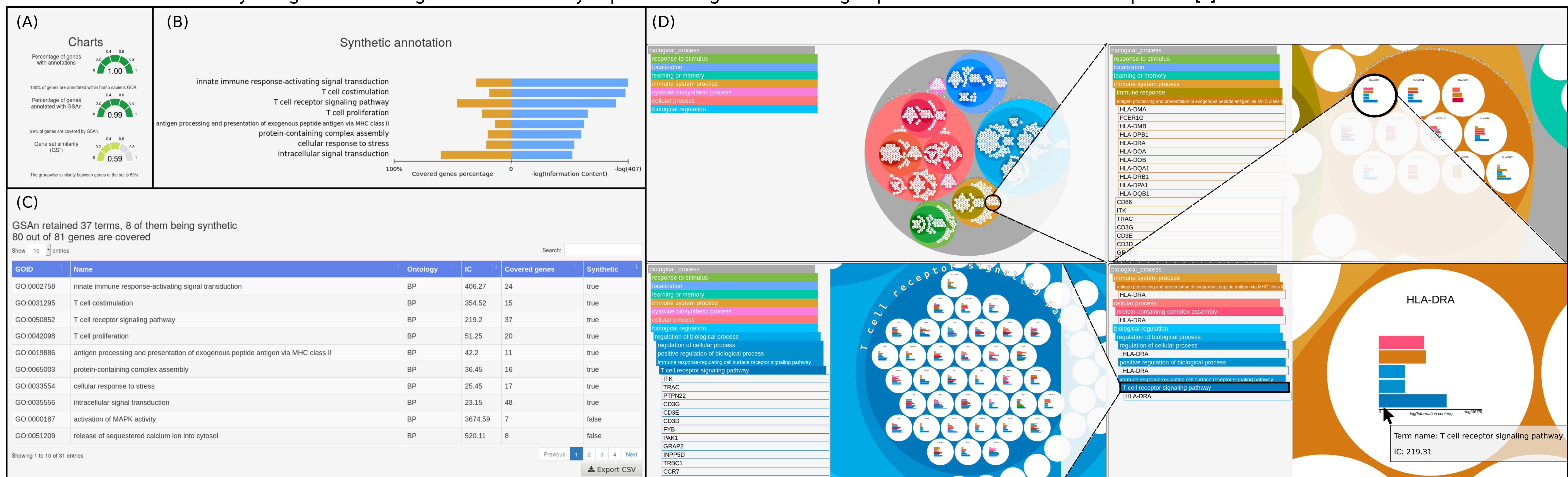
- These statistical methods tend to focus on the most studied genes and provide gene set annotation results that cover a limited number of annotated genes [1, 2, 3].
- Visualization facilities often suffer from a lack of capacity to perform multi-scale analyses which may help users while interpreting their results.

3 - Method



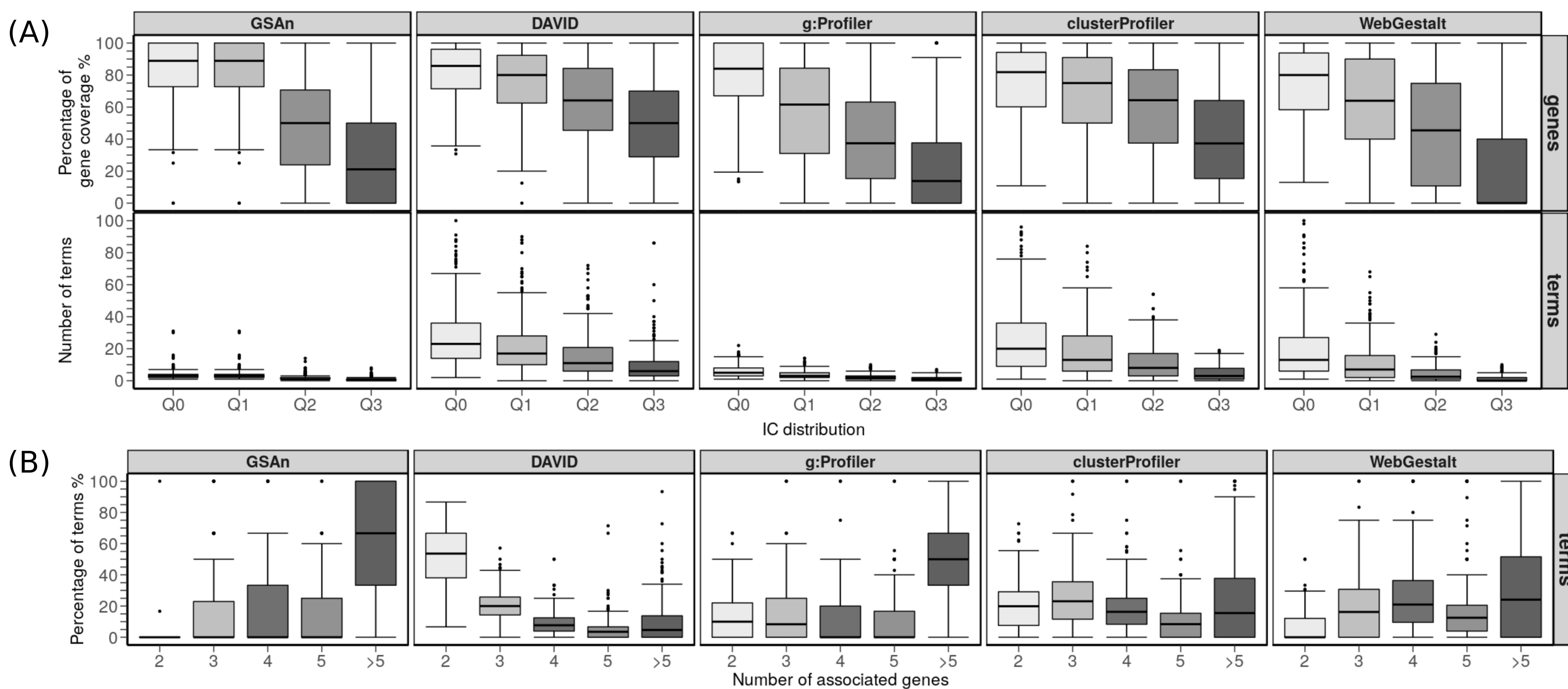
4 - GSAN results using an immunological gene set

- We used GSAN to analyze a gene set of 81 genes annotated by experts as "regulation of antigen presentation and immune response" [4].



5 - Comparison with other gene set annotation tools

Four enrichment analysis were compared in order to observe their impact in the annotation of gene set by using 346 gene sets provided by [4]. (A) We used four thresholds (from Q0 to Q3) corresponding to the quartiles of the Information Content (IC) distribution (0.0, 18.4, 44.4 and 155.3, respectively) to filter out the results of each tool. Each threshold filters out the terms with an IC below their value. (B) To further investigate this comparison, we analyzed the percentage of terms according to the number of genes that are annotated by these terms (2, 3, 4, 5, and more than 5 genes).



Availability

GSAN web service is available at <https://gsan.labri.fr>

References

- Tomczak A. et al. Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. *Sci. Rep.*, 2018
- Haynes W.A. et al. Gene annotation bias impedes biomedical research. *Sci. Rep.*, 2018
- Bleazard T. et al. Bias in microRNA functional enrichment analysis. *Bioinformatics*, 2015
- Li S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat. Immunol.*, 2013
- Ayllón-Benítez A. et al. A new method for evaluating the impacts of semantic similarity measures on the annotation of gene sets. *PLoS ONE*, 2018
- Ayllón-Benítez A. et al. Deciphering gene sets annotations with ontology based visualization. 21st international conference Information Visualization 2017

Contact

aaron.ayllon-benitez@u-bordeaux.fr
ayllon-benitez.eu